

Test-Retest Reproducibility of Heart Rate Recovery After Treadmill Exercise

Barbara P. Yawn, MD, MSc¹

K. Afzal Ammar, MD²

Randal Thomas, MD, MS³

Peter C. Wollan, PhD¹

¹Department of Research, Olmsted Medical Center, Rochester, Minn

²Department of Internal Medicine, Olmsted Medical Center, Rochester, Minn

³Department of Cardiovascular Diseases and Internal Medicine, Mayo Clinic and Foundation, Rochester, Minn

ABSTRACT

BACKGROUND Slowed heart rate recovery (HRR) of less than 12 beats per minute in the first minute after an exercise stress test has been suggested as a useful addition to the criteria currently used to assess exercise stress test results. Although HRR has been tested in large populations, the short-term test-retest stability (reproducibility) of abnormal HRR for an individual has not been assessed.

METHODS The study was a retrospective comparison of medical record information using a community-practice-based sample of 90 patients undergoing 2 exercise stress tests separated by 18 weeks or less. Concordance of abnormal HRR results on the first and second stress tests were assessed for individual patients using definitions of abnormal HRR from the medical literature.

RESULTS Individual patient's HRR was markedly variable from the first to second stress test. In this sample, no definition of abnormal HRR provided more than 55% concordance between results from the first and second stress tests.

CONCLUSION These preliminary data suggest that HRR appears to have limited short-term test-retest stability or reproducibility and therefore might not be a reliable addition to current results of exercise stress tests.

Ann Fam Med 2003;1:236-241. DOI: 10.1370/afm.37.

INTRODUCTION

Heart disease is the second leading diagnosis in patients visiting the offices of adult generalist physicians (family physicians and general internists).¹ More than 75% of general internists and 13% of family physicians perform and interpret exercise stress tests in their offices.^{1,2} Many other generalist physicians must help patients understand the results of exercise stress tests performed by cardiologists and integrate the test results into their management plans. Currently several indicators, such as symptoms and ST segment elevation and T wave inversion on electrocardiogram (ECG) tracings, are included in stress test interpretation to determine which patients are at increased risk for cardiac mortality.³

Heart rate recovery (HRR), the ability to slow the heart rate after an exercise stress test, has been suggested as a measure of chronotropic competence⁴⁻⁹ and as a useful addition to current criteria for predicting cardiac and all-cause mortality from an exercise stress test.^{4-8,10-15} Recent work has focused on the decrease in heart rate within a fixed period after termination of the stress test.^{4-7,10-15} The definition of abnormal has varied both in the time frame and in the magnitude of the decrease, but the most commonly used value defines abnormal HRR as a decrease of fewer than 12 beats per minute within the first minute.¹¹ In both community-based and cardiology clinic populations, slower HRR (slower decrease in beats per minute in the first 1 or 2 minutes after exercise) has been associated with all-cause mortality, but it has not been linked reproducibly with cardiac-related mortality,^{5-7,11-13,15,16} nor has it been an independent risk factor for abnormal angiography findings.^{4,14}

Conflicts of interest: none reported

CORRESPONDING AUTHOR

Barbara P. Yawn, MD, MSc
Department of Research
Olmsted Medical Center
210 Ninth St SE
Rochester, MN 55904
yawnx002@tc.umn.edu

Although HRR is suggested as a useful addition to exercise stress test results,¹⁴ the specificity and sensitivity of an abnormal test for an individual patient have not been published,^{4, 11-13} and no agreement has been reached regarding a single numerical value to define an abnormal HRR.¹¹⁻¹⁵ Furthermore, no studies have validated the short-term reproducibility (test-retest reliability) of an abnormal HRR. For HRR to be a reliable predictor of mortality risk in the individual patient, its measurement should be stable (in the individual) for at least a short period (4 to 6 months) when no intervening cardiac events or therapeutic interventions (pharmaceutical or surgical) have occurred. If HRR is not found to be stable under these circumstances, its role as a prognostic tool appears limited. This preliminary study assesses the short-term reproducibility of HRR as a first step in evaluating its clinical utility.

METHODS

After receiving approval from the Olmsted Medical Center Institutional Review Board, data were retrospectively analyzed from exercise stress tests that were ordered for clinical evaluation of symptomatic and asymptomatic men and women aged between 30 and 80 years. Data were obtained from each patient's medical records and from the original printouts from the exercise stress test.

Sample

The subjects were residents of southeastern Minnesota who were seen in the Olmsted Medical Center Department of Internal Medicine for exercise stress testing between January 1, 1998, and January 1, 2002. Patients included in the study underwent a second exercise stress test within 18 weeks. Patients who had a second test after an intervening event or therapeutic intervention were excluded because these events might be expected to affect the test results and the prediction of the patient's 5-year mortality. In all included cases the ECG findings on the first test were interpreted as being equivocal or not diagnostic for inducible myocardial ischemia, which led to further evaluation with a second test augmented by either echocardiography or nuclear imaging. Patients who had documented noncardiac comorbidities that left them unable to exercise sufficiently to produce symptoms (eg, chronic obstructive pulmonary disease, arthritis) were excluded.

The population of southeastern Minnesota is predominantly (more than 85%) white and non-Hispanic. The region is primarily rural with one city of about 90,000 residents, Rochester, Minn. The community practice (Olmsted Medical Center) has a main office in Rochester that provides all exercise stress tests for the main clinic and its 11 satellite clinics throughout

southeastern Minnesota. All tests were supervised by 1 of 8 general internists working at the Rochester site.

Identification of the Sample

Using billing and administrative data, all patients with 2 exercise stress tests completed within any 18-week period from January 1, 1998, to January 1, 2002, were evaluated as potential subjects for the study. Their medical records were reviewed to assess the occurrence of an intervening event. Of the 158 potential subjects, 64 were excluded because of a documented occurrence of an event between the 2 tests that might affect cardiac chronotropic competence, including acute myocardial infarction, placement of a cardiac pacemaker, congestive heart failure, a hospital admission for unstable angina, cardiac arrest, stroke or transient ischemic attack, percutaneous coronary angioplasty, coronary artery bypass graft, use of an intravenous thrombolytic agent, or the initiation of digoxin or β -blocking agents. Patients were also excluded ($n = 4$) if they had refused the general research authorization required by Minnesota statute.

Data Collection

Using all medical records from the Olmsted Medical Center (office, emergency department, and hospital), an experienced nurse abstractor reviewed the records to determine the indication for the exercise stress test, demographic information, physician performing the test, maximal heart rate, heart rate at 1, 2, and 4 minutes post-test, the type of recovery phase used, and the presence of symptoms. Data on the exercise stress tests were collected from the printed and electronic databases kept for all tests. Stress test heart rates were taken from the computerized reports, which are based on the last 5 R-R intervals prior to the defined times to prevent spurious results based on normal variants, such as sinus arrhythmias.

All exercise stress tests were treadmill stress tests using the Bruce protocol and were symptom limited or pushed to 90% of maximal heart rate in the absence of symptoms. For all non-echocardiogram-augmented exercise stress tests, 5 internists used a passive recovery phase (immediately lying down at exercise completion), 2 internists used an active recovery phase (slow walk for 1 minute), and the remaining internist had the patient continue a slow walk while blood pressure was recorded (30 seconds) and the patient was then seated. A passive recovery phase was used for all echocardiogram-augmented stress tests.

Data Analysis

Summary statistics were completed for the population. HRR was calculated as the peak heart rate (highest heart rate during the stress test) minus the heart rate at 1, 2, and 4 minutes after testing. Paired HRR values

were graphed for visual assessment using different symbols for men and women. Heart rates from the first and second test were compared for each individual using Bland-Altman plots.¹⁷ These analyses were done for all subjects and then for the subset of 44 patients (22 men and 22 women) who had passive recovery phases on each of the 2 tests. Both analyses are included because in the community practice a single standard for HRR applicable to all testing regardless of the manner of recovery would be most useful. If a single standard is not possible, however, then it is necessary to look at reproducibility using a single type of recovery method.

Reproducibility was quantified in 2 ways: first, as the number of patients with an abnormal HRR on both tests divided by the number with an abnormal HRR on the first test and converted to a percentage; and second, as the standard deviation of the difference between first and second HRR. Association of reproducibility with other factors (age, sex, time between tests, active vs passive recovery, and peak heart rate) was assessed visually with plots and formally tested by analysis of variance or linear regression on the absolute value of the difference between the first and second HRR (modified Levene test).

RESULTS

The study sample consisted of 90 subjects (49 men and 41 women), which accounts for 13% of all exercise stress testing done in this setting during the study period. Ages ranged from 31 to 76 years, with a similar ($P = .69$) mean age for men (53.4 years) and women (54.3 years). The mean time between test 1 and 2 was 29 days and was not different for men and women ($P = .75$). Whereas all 90 subjects had equivocal or ECG changes that were not diagnostic on the initial test, 40 (44%) had negative results on the test aug-

mented by echocardiography or nuclear imaging, 47 (52%) had results consistent with coronary artery disease, and 3 (3%) had results that remained equivocal.

The decrease in heart rate from peak heart rate to 1 minute into the recovery is displayed in Figure 1 for all subjects ($N = 90$) and in Figure 2 for those with passive

Figure 1. One-minute heart rate recovery (HRR): test 1 vs test 2 for entire sample (N=90).

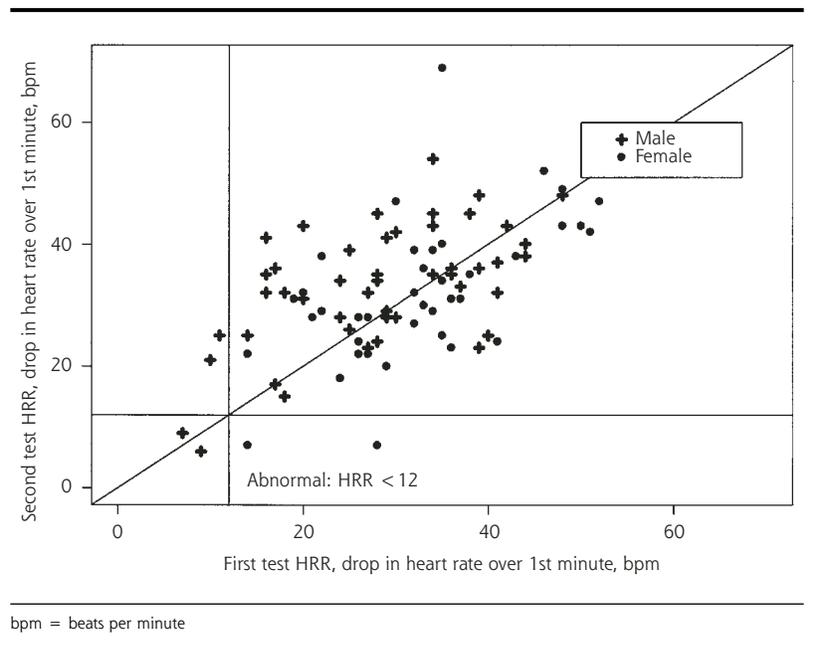
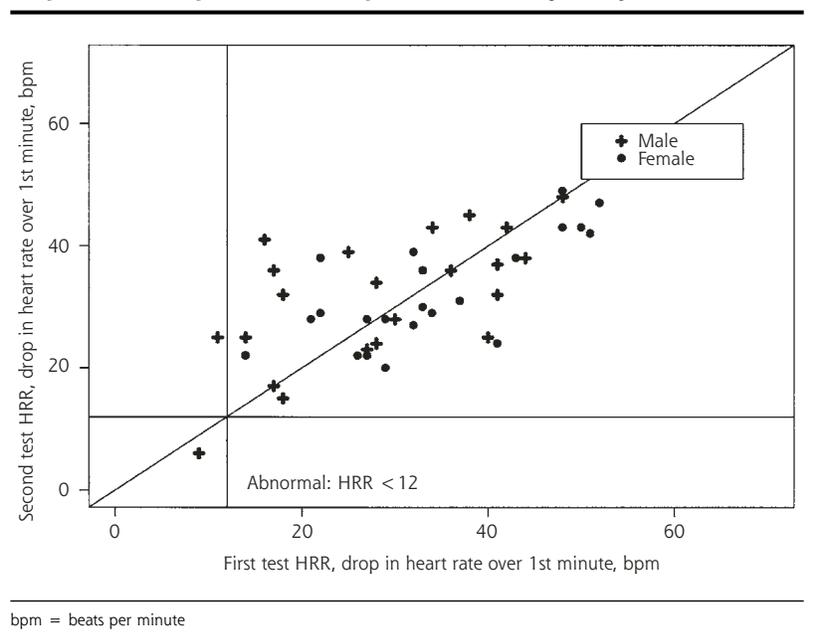


Figure 2. One-minute heart rate recovery (HRR): test 1 vs test 2 for only those with passive recovery on both tests (n=44).



recovery in both tests ($n = 44$). When all subjects are included, using the criterion of abnormal HRR as a decrease of fewer than 12 beats per minute¹¹ between peak heart rate and heart rate 1 minute into recovery, 6 patients had abnormal results after at least 1 of the stress tests. Two had abnormal results on both tests, 2 had abnormal results on the first test only, and 2 had abnormal results on the second test only, providing a 50% reproducibility. Using the subset of only those with passive recovery on both tests ($n = 44$), 1 had abnormal results on both tests, 1 had abnormal results on the first test only, and none had abnormal results on the second test, providing 50% reproducibility. The use of other criteria for abnormal (a decrease of 18 beats per minute or less at 1 minute¹³ and a decrease of fewer than 21 beats per minute at 2 minutes⁶) did not raise the percentage of reproducibility to higher than 55%.

Figures 1 and 2 show that the recovery rates after the first and second tests are strongly correlated. At issue is whether the measurements are sufficiently consistent for clinical diagnostic purposes. One way to address this question is to compare normal-abnormal classifications, as above. Because we had very few abnormal results, a more informative approach is to display the variability in recovery rates graphically: the Bland-Altman plot of the difference between the 2 HRR values vs the mean is displayed in Figure 3. The standard deviation of the difference in HRR for the 2 tests (9.5 for men and 9.8 for women) is large compared with the cutoff of 12 beats per minute used to

define an abnormal 1-minute HRR. Limiting the analysis to those with passive recovery did not improve the results. Results were similar for 2- and 4-minute recovery times.

HRR was slightly more reproducible when the decrease in beats per minute from peak heart rate to the heart rate at 1-minute posttest was greater, but the trend was not statistically significant ($P = .07$). Reproducibility of HRR was not associated with age of the subject ($P = .63$), sex of the subject ($P = .55$), or the time between the 2 tests ($P = .59$). There was no difference in variability of HRR between active recovery and passive recovery methods ($P = .88$). There was greater variability in instances where 1 of the 2 tests used active and the other passive recovery than there was when both tests had the same recovery method, whether active or passive, but the difference was not statistically significant ($P = .13$).

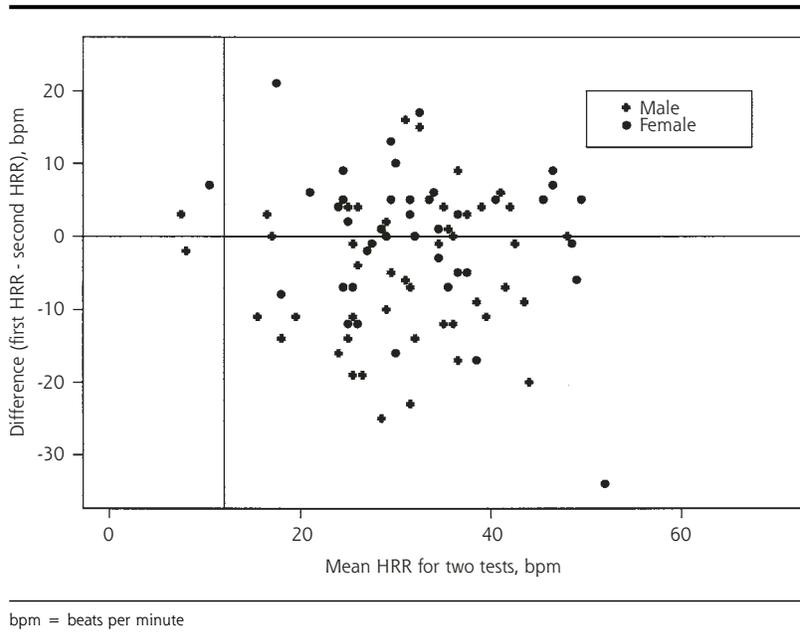
DISCUSSION

HRR after exercise stress testing does not have short-term test-retest reproducibility in this community-practice-based sample of men and women. The measure is not reproducible using any of the published criteria for abnormal HRR. The percentage of the population having at least 1 test with an abnormal HRR varies from 2% to 46% depending on the definition that is selected from the medical literature, and reproducibility of abnormal is never greater than 55%. This study is the

first to assess the test-retest reproducibility of HRR and calls into question the value of the measure in clinical practice.

Other researchers have found similar problems in test-retest reproducibility of cardiovascular responses to different types of stress.^{10,18-20} Christenfield et al¹⁸ found that healthy young adults given 3 stressors (walking, mental arithmetic, and public speaking) did not have reproducible cardiovascular recovery measured as both blood pressure and HRR. Similarly, data from the Bootsma et al study of tilt table tests in healthy young volunteers²¹ suggest that a person's autonomic dynamics vary within even a short period. Other studies have reported significant variability in peak heart rates during repeated exercise stress testing.^{19,20}

Figure 3. Difference in 1-minute heart rate recovery (HRR) vs average heart rate recovery (N=90).



The use of HRR as a population-based prognostic marker of mortality has been validated across different populations.^{4,12-14} Previously published studies differ by the groups of subjects that were excluded, the use of heart rate or symptom-limited protocols, and an active or a passive recovery phase.^{5,11-14} All studies found that statistical evaluation of the data could produce a measure of HRR which was associated with higher 5-year all-cause mortality for the population; however, not only were there different numerical cutoff values for abnormal HRR, there were also different time points for the measurement. These marked variations in the published definitions also suggest a lack of reproducibility of a specific HRR measure.^{14,22,23}

Our results indicate that HRR is not reproducible when measured in this natural experiment. We did not have tight control over either the exercise stress test procedure or the patient population, which is typical of real-world practice variations. It is possible that HRR could be shown to be reproducible under suitable restrictions. Future studies will need to specify the exact protocol and recovery phase, as well as the cutoff for abnormal HRR. These preliminary data highlight the need for prospective trials that include test-retest reliability assessment.

Potential Limitations

Our study has all of the limitations of a retrospective natural experiment in a community setting. The tests were separated by weeks rather than hours. The physicians all used symptom-limited protocols for the stress tests but different types of recovery phases. Limiting the assessment of reproducibility to those with patients who had passive recovery phases in both tests, however, did not improve the test-retest reliability or the concordance of abnormal HRR between the 2 tests.

The population studied was not a random sample of the community residents but rather was a sample of patients from 1 medical group with clinical indications for stress testing. All patients had a second test ordered to clarify nondiagnostic or equivocal ECG information on the exercise stress test results. These patients did not appear to be sicker than the usual patients undergoing diagnostic stress testing, as 40% of them were found to have negative results on a second test augmented by echocardiography or nuclear scan. The careful removal of all patients with any documented intervening event should also remove patients with sudden changes in their cardiac status and their risk for 5-year mortality. Most patients in primary care require only a single exercise stress test for evaluation, and it is possible that HRR results in such patients might be more reproducible than in our group of patients that had 2 tests performed. The patients studied are 98% white and

non-Hispanic. It is possible that HRR might be more reproducible in other racial populations. The sample was almost equally divided between men and women, facilitating the ability to show a lack of reproducibility in both sexes. Like several of the previous studies of HRR, we do not have anatomical proof of heart disease in any of the subjects. Proof of heart disease should not be a limitation, because HRR is suggested to be an independent risk factor of all-cause mortality.

The study was intended as a first look to find out whether HRR satisfies the basic minimal requirement that a repeated test in a patient with no change in clinical status should have the same result. The lack of reproducibility in our data could be related to the variability in the test procedure, differences within this subgroup of patients, or inherent variability in the process being reviewed. This study was proposed by a generalist physician in full-time practice who suspected the latter. The work was carried out in collaboration with a community-based department of research and highlights the importance of natural experiments in the real world of community practices to augment studies done in specialized populations or by special study personnel.

CONCLUSION

In this community sample the lack of test-retest reproducibility in individual patients suggests that HRR is not ready for translation into general clinical practice to augment the interpretation of exercise stress tests. The study of chronotropic efficiency is important and should be continued with prospective community-based studies to confirm or refute these preliminary findings.

To read commentaries or to post a response to this article, see the online version at <http://www.annfammed.org/cgi/content/full/1/4/236>.

Key words: Exercise tests; electrocardiography; heart rate; prognosis; cardiology; diagnostic techniques, cardiovascular; cardiovascular system

Submitted November 5, 2002; submitted, revised, January 30, 2003; accepted February 17, 2003.

Funding support: This research was funded by a grant from the Agency for Healthcare Policy and Research, grant No. R01 HS10239.

References

1. Moskowitz D, Kozzumi L, eds. *2000 Health Care Almanac & Yearbook*. New York, NY: Faulkner & Gray; 2000.
2. American Academy of Family Physicians. *Facts about Family Practice*. Kansas City, Mo: American Academy of Family Physicians; 1998.
3. Ellestad MH, Wan MK. Predictive implications of stress testing: follow-up of 2700 subjects after maximum treadmill stress testing. *Circulation*. 1975;51:363-369.

4. Desai MY, De la Pena-Almaguer E, Manning F. Abnormal heart rate recovery after exercise as a reflection of an abnormal chronotropic response. *Am J Cardiol.* 2001;87:1164-1169.
5. Lauer MS, Okin PM, Larson MG, Evans JC, Levy D. Impaired heart rate response to graded exercise. Prognostic implications of chronotropic incompetence in the Framingham Heart Study. *Circulation.* 1996;93:1520-1526.
6. Lauer MS, Francis GS, Okin PM, Pashkow FJ, Snader CE, Marwick TH. Impaired chronotropic response to exercise stress testing as a predictor of mortality. *JAMA.* 1999;281:524-529.
7. Dresing TJ, Blackstone EH, Pashkow FJ, Snader CE, Marwick TH, Lauer MS. Usefulness of impaired chronotropic response to exercise as a predictor of mortality, independent of the severity of coronary artery disease. *Am J Cardiol.* 2000;86:602-609.
8. Wilkoff BL, Miller RE. Exercise testing for chronotropic assessment. *Cardiol Clin.* 1992;10:705-717.
9. Hammermeister KE, DeRouen TA, Dodge T. Variables predictive of survival in patients with coronary disease. Selection by univariate and multivariate analyses from the clinical, electrocardiographic, exercise, arteriographic, and quantitative angiographic evaluations. *Circulation.* 1979;59:421-430.
10. Hammond HK, Froelicher VF. Normal and abnormal heart rate responses to exercise. *Prog Cardiovasc Dis.* 1985;27:271-296.
11. Cole CR, Blackstone EH, Pashkow FJ, Snader CE, Lauer MS. Heart-rate recovery immediately after exercise as a predictor of mortality. *N Engl J Med.* 1999;341:1351-1357.
12. Cole CR, Foody JM, Blackstone EH, Lauer MS. Heart rate recovery after submaximal exercise testing as a predictor of mortality in a cardiovascularly healthy cohort. *Ann Intern Med.* 2000;132:552-555.
13. Nishime EO, Cole CR, Blackstone EH, Pashkow FJ, Lauer MS. Heart rate recovery and treadmill exercise score as predictors of mortality in patients referred for exercise ECG. *JAMA.* 2000;284:1392-1398.
14. Shetler K, Marcus R, Froelicher VF, et al. Heart rate recovery: validation and methodologic issues. *J Am Coll Cardiol.* 2001;38:1980-1987.
15. Watanabe J, Thamilarasan M, Blackstone EH, Thomas JD, Lauer MS. Heart rate recovery immediately after treadmill exercise and left ventricular systolic dysfunction as predictors of mortality: the case of stress echocardiography. *Circulation.* 2001;104:1911-1916.
16. Diaz LA, Brunken RC, Blackstone EH, Snader CE, Lauer MS. Independent contribution of myocardial perfusion defects to exercise capacity and heart rate recovery for prediction of all-cause mortality in patients with known or suspected coronary heart disease. *J Am Coll Cardiol.* 2001;37:1558-1564.
17. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* 1986;1:307-310.
18. Christenfeld N, Glynn LM, Gerin W. On the reliable assessment of cardiovascular recovery: an application of curve-fitting techniques. *Psychophysiology.* 2000;37:543-550.
19. Linden RJ, Mary DA. Limitations and reliability of exercise electrocardiography tests in coronary heart disease. *Cardiovasc Res.* 1982;16:675-710.
20. Tonino RP, Driscoll PA. Reliability of maximal and submaximal parameters of treadmill testing for the measurement of physical training in older persons. *J Gerontol.* 1988;43:M101-M104.
21. Bootsma M, Swenne CA, Bruschke AV. Heart rate variability during repeated incremental head-up tilt discloses time dependence of individual autonomic dynamics. *Clin Cardiol.* 1996;19:62-68.
22. Goldstein RE, Holmboe ES. Prognostic indicators for coronary artery disease: ready for the bedside? *JAMA.* 1999;281:565-566.
23. Gammethaler S. Recovery of heart rate after exercise. *N Engl J Med.* 2000;342:662.